3H

УДК 004.415:004.822

С.П. Некрашевич
Государственный университет информатики и искусственного интеллекта,
г. Донецк, Украина
netrovich@gmail.com

Построение модели онтологии интеллектуальной системы мониторинга учебного процесса дистанционного образования

В статье рассматривается построение онтологии интеллектуальной системы мониторинга учебного процесса дистанционного образования (ДО). Приводятся различные модели онтологий, аргументируется их использование в виде онтологической системы. Дается формальное описание онтологии интеллектуальной системы предметной области ДО на основе кортежей множеств. В качестве концептуального моделирования используются диаграммы UML, расширенные дополнительными типами отношений.

Компьютерная система дистанционного образования (ДО) предназначена для адаптации процесса обучения к индивидуальным характеристикам обучаемых, она освобождает вовлеченных в учебный процесс преподавателей, студентов и администрацию от ряда трудоемких и рутинных операций по представлению учебной информации и контролю знаний, что способствует разработке объективных методов контроля знаний и облегчению накопления учебно-методического опыта.

Разрабатываемая интеллектуальная система, использующая шифрование управляющей информации маркерами расширенной реальности ARGET [1], предоставляет измеримые метрики процесса дистанционного образования и осуществляет мониторинг учебного процесса, а также его операционный анализ и прогнозирование. Сокращение длины шифрограммы основывается на использовании специального, перестраиваемого системой словаря смыслов управляющей информации. Словарь строится интеллектуальной системой на основании онтологической модели учебного процесса.

Поэтому центральное место в системе занимают онтологии. Они предоставляют концептуальную модель задач и предметной области, определяют реализацию бизнес-логики системы. Учебный процесс, его структура и бизнес-процессы описываются в терминах связанных знаний.

Онтология является спецификацией концептуализации предметной области [2]. Это формальное и декларативное представление, которое включает словарь понятий и соответствующих им терминов предметной области, а также логические выражения, которые описывают множество отношений между понятиями. Для описания отношений в онтологиях используются различные формальные модели и языки, существующие в искусственном интеллекте – предикаты, продукции, фреймы, семантические сети и др. Термин «онтология» является синонимом представления знаний.

Формально онтология определяется как тройка множеств [3, с. 286-291]:

$$O = \langle X, R, F \rangle$$

X – множество концептов (понятий, терминов) предметной области, которую представляет онтология O;
R – множество отношений между концептами заданной предметной области;
F – множество функций интерпретации (аксиоматизация), заданных на концептах и/или отношениях онтологии O.

Предметная область ДО требует использования различных моделей онтологий, представленных в табл. 1.

Таблица 1 – Классификация моделей онтологии

Тип модели	Формальное определение
Словарь ПО	$<X, \varnothing, \varnothing>$
Пассивный словарь ПО	$<X, \varnothing, F>$, $X = X_1 \cup X_2$, $X_1 \cap X_2 = \varnothing$, $F = \{:=\}$ X_1 – множество интерпретируемых терминов X_2 – множество интерпретирующих терминов
Активный словарь ПО	$<X, \varnothing, F>$, $X = X_1 \cup X_2$, $X_1 \cap X_2 = \varnothing$ $\exists (x \in X_1, y^1, y^2, ..., y^k \in X_2), x = f(y^1, y^2, ..., y^k), f \in F$
Таксономия понятий ПО	$<X, R, \varnothing>$, R = {is-a}
Сетевая структура	$<X, R, \varnothing>$, R = {is-a, part-of, отношения ПО}
Расширяемая сетевая структура	$<X, R, F>$, R = {is-a, part-of, отношения ПО, F(R)}
Расширяемая онтология	$<X, R, F>$, F(X), F(R)

Словарь предметной области требует наличия только концептов X, в то время как таксономия понятий требует дополнительного указания отношений между этими концептами в виде отношения is-a для моделирования наследования концептов.

Для разделения понятий предметной области на синонимы, антонимы и другие классы дескриптивной логики вводятся отношения нескольких типов, соответствующие этим классам разделения (например, синоним, антоним, отлично-от).

Структурная (композиционная) схема, являющаяся подтипом сетевой структуры, использует только один тип отношений part-of (является-частью) или обратное ему contains (содержит). Пример структурной схемы ДО приведен на рис. 1.

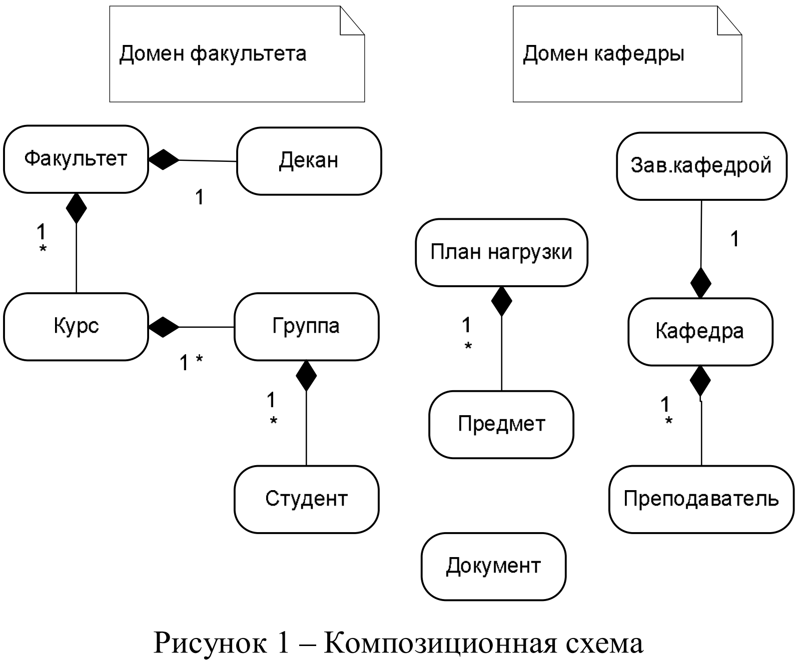


Рисунок 1 – Композиционная схема

Таким образом, использование отношений между концептами занимает важное место в модели онтологии. Многообразие различных типов отношений позволяет моделировать не только информационную структуру, но и операционную семантику понятий предметной области ДО.

Предметная область учебного процесса ДО содержит достаточно большое количество связанных понятий и их экземпляров. Концептуальное моделирование и проектирование модели онтологии интеллектуальной системы этой предметной области требует применения многоуровневой системы онтологий различных типов. Такая система может быть определена общей онтологией без учета доменов, которые моделируются отдельно.

Использование множества концептов целесообразно в виде онтологической системы, разделяющей уровни (домены) и выделяющие компоненты сильной связности в виде отдельных онтологий различных типов.

Формальное определение онтологической системы

$$S = <O^{meta}, \{O^{domain}\}, I^{f}>$$

O^{meta} – метаонтология <X, R, ∅>;
$\{O^{domain}\}$ – множество предметных онтологий <X, R, F> и онтологий задач ПО <X, ∅, ∅ >;
I^{f} – машина вывода, ассоциированная с S.

Работа машины вывода характеризуется описанием исходной ситуации, определением целевой ситуации и выводом на сети посредством распространения волн активации от узлов исходной ситуации, использующих свойства отношений, с ними связанные.

Формальную модель результирующей онтологии ИС предметной области ДО можно представить в виде кортежей множеств.

O = <X, R, F>
X = <N, A, E>
R = <N, {X}i, i=1..*>
E = <N, P, X>
A = <N, D, C, X>
P = <N, A, V>
D = <N, метаописание, F>
V = <D, C, значение, F>
N = <название на языке L>
L = <язык>
F = <функции интерпретации>
C = <F>

Наиболее важным элементом модели онтологии является понятие концепта (concept). Каждому концепту из множества $X = \{x_i \mid i=1..N_X\}$ ставится в соответствие тройка из множеств названий, атрибутов и экземпляров концептов. Наличие экземпляров E в этом кортеже позволяет определять концепт как отдельный класс через множество составляющих его экземпляров.

Следующим важным элементом модели, как было показано выше, является отношение между концептами (relation). Каждому отношению из множества $R = \{r_i \mid i=1..N_R\}$ ставится в соответствие пара множеств названий и связей между концептами. Данная модель позволяет использовать произвольное число концептов, участвующих в связи, начиная с традиционных и повсеместно используемых бинарных отношений,

заканчивая отношениями более высокой кратности. Отдельно следует отметить наличие отношений кратности 1, при котором концепт находится в отношении к самому себе (внутреннее замыкание), использование такого отношения позволяет моделировать порядок, выстраивающий иерархию экземпляров одного концепта, например, зависимости документов, иерархии сотрудников и др. Отношения высокой кратности можно свести (редуцировать) к бинарным и использовать их в модели, однако это увеличивает количество отношений и снижает уровень абстракции модели в целом.

На рис. 2 изображены отношения различных типов, включающие наследование и ассоциации. Не показаны отношения высокой кратности между преподавателем, студентом и предметом для описания тернарной ассоциации.

В классической сетевой семантической модели отношения можно рассматривать как подмножество концептов на основании их принадлежности этому множеству в виде самостоятельных понятий предметной области, связывающих концепты [4].

Атрибуты концептов также рассматриваются как значения отношения has (имеет атрибут) между концептом и самим атрибутом, однако в данной модели все атрибуты концептов выделяются в отдельное множество, ответственное за домен атрибутов.

Аналогичным образом экземпляры концепта можно рассматривать как значения отношения instance-of между концептом и самим объектом, однако в данной модели экземпляры выделяются в отдельное множество, ответственное за домен объектов.

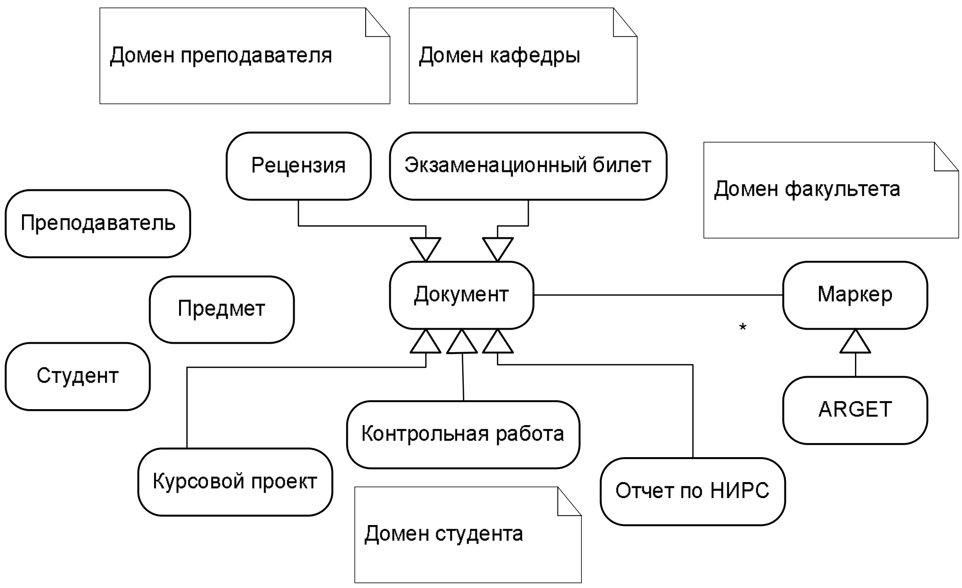


Рисунок 2 – Тезаурус документов

Атрибуты (attribute) концептов $A = \{a_i \mid i=1..N_A\}$ описывают структуру и домен свойств каждого концепта и определяются множествами названий, типов данных (datatype) и ограничений. Использование X в кортеже атрибута A позволяет задавать домен применения атрибутов среди концептов, например, наличие атрибутов даты или маркера регистрации у некоторых концептов позволяет выделить их в отдельный класс понятия (все, что можно регистрировать).

Семантика типа данных (datatype) $D = \{d_i \mid i=1..N_D\}$ определяется в метаонтологии и описывается набором применимых операций. Для анализа такой семантики и применения в информационной модели необходимо использовать интерпретирующие правила F. Существует достаточно много подходов к моделированию и семантическому описанию системы типов данных, включая промышленные стандарты описания схем данных, однако их рассмотрение выходит за рамки статьи.

Каждому из экземпляров (entity) концептов $E = \{e_i \mid i=1..N_E\}$ ставится в соответствие кортеж из множеств названий, свойств P экземляров и концептов X. Использование X в кортеже позволяет моделировать роли, в которых может выступать каждый отдельный экземпляр. Например, конкретный документ, используемый в процессе ДО, может рассматриваться как экземпляр различных классов (текст, гипертекст, множество изображений листов, озвучка текста на естественном языке и др.), что соответствует как множественному наследованию в объектно-ориентированном подходе, так и применению более гибкой системы реализации объектами интерфейсов, которые широко используются в компонентном программировании и технологиях Corba, COM, Java RMI и др.

Каждому из свойств (property) экземпляров $P = \{p_i \mid i=1..N_P\}$ ставится в соответствие кортеж из множеств атрибутов A и данных V, соответствующих значениям этих атрибутов.

Данные из множества $V = \{v_i \mid i=1..N_V\}$ являются терминальными (конечными) элементами модели, их конкретные значения не хранятся непосредственно в системе, а находятся в специализированном хранилище данных для удобства аналитической обработки, источником данных также может служить F.

С точки зрения логики более высокого порядка, данные являются экземплярами типов данных и должны быть представлены подмножеством экземпляров E, в то время как типы – отдельными концептами X. Эта логика сохраняется только для типов данных. На практике часто данные хранятся отдельно от схемы, их описывающей, примером являются схемы баз данных или схемы описания данных, представленных в формате XML.

Ограничение (constraint) $C = \{c_i \mid i=1..N_C\}$ является исключением из общего правила определения типа D и используется в качестве механизма ограничения не только значений, но и ограничения мощности используемых множеств. Задание ограничения исключает создание множеств элементов, имеющих незначительные различия. Вместо этого создается общее множество элементов и накладываются ограничения на его отдельные элементы.

Каждый элемент модели при своем создании получает уникальный идентификатор id, который связан с элементом все время его существования и не меняется при изменении значения элемента. Идентификатор является частью значения элемента и обеспечивает его идентификацию в интеллектуальной системе. В качестве идентификатора также может использоваться элемент названия N, однако это приводит к избыточности модели и в случае наличия повторений в множестве N могут потребоваться дополнительные действия (вывод) для идентификации конкретного элемента по его имени.

Использование нескольких названий на различных языках позволяет моделировать синонимы без введения дополнительных концептов, а также обеспечивает интероперабельность в среде с несколькими языками. Например, используемый в интеллектуальной системе документ может быть закодирован с использованием названий на одном языке, а проинтерпретирован на другом языке с сохранением исходной семантики.

Множество языков $L = \{lang_i \mid i=1..N_L\}$ представляет основу для понятий лингвистической онтологии [5], а именно, все названия и лексически зависимые определения, на которых представлена онтология описания понятий, их экземпляров, атрибутов и отношений как на естественном языке, так и с использованием искусственных нотаций, на которых формально представлены тесты модели. Использование нескольких естественных языков позволяет разделять домены, например, государственного языка и делопроизводства (украинский язык), язык повседневного общения (русский язык), домен международных связей (английский язык).

Для облегчения визуализации проводится моделирование предметной области на языке UML [6], оно заключается в описании концептов и отношений с учетом различных доменов.

Использование нотации UML оправдано ввиду наличия удобных средств моделирования, концептуального в том числе, и наличия удобных способов представления статической структуры и динамического поведения объектов.

Однако, учитывая то, что UML ориентирован в первую очередь на разработку программного обеспечения на основе объектно-ориентированного подхода, необходима некоторая адаптация его применения для онтологического моделирования, поэтому вносятся некоторые расширения:

– состояние и поведение объектов на данном этапе концептуального моделирования не указываются, поскольку они больше важны для построения информационной модели и архитектуры реализации и будут доопределяться на более поздних этапах;
– расширены типы отношений между понятиями – добавлены некоторые отношения предметной области;
– использованы несколько отношений между классами;
– использованы внутренние отношения (отношение порядка, выстраивающее иерархию подчинения в коллекции).

Результирующая модель онтологии может быть представлена семантической сетью, которая интерпретирует онтологию с точки зрения элементарных составляющих на низком уровне абстракции. После доопределения атрибутов концептов может быть построена и реализована информационная модель, например, в виде реляционной базы данных.

Выводы

Предложенная формальная модель онтологии интеллектуальной системы мониторинга учебного процесса дистанционного образования является первичным этапом в разработке интеллектуальной системы, требуется дальнейшая разработка вопросов проектирования и наполнения онтологии по предложенной модели, ее описание в формате хранилища данных и представление на формальном языке (например, OWL), а также определение и использование функций интерпретации F и применение их в задачах бизнес-процессов предметной области ДО.

Литература

1. Гудаев О.А. Синтез и анализ предложений графического языка передачи сообщений в мобильных робототехнических системах с элементами расширенной реальности (ARGET) / О.А. Гудаев // Искусственный интеллект. – 2006. – № 2. – С. 467-498.
2. Gruber T.R. A translation approach to portable ontologies / T.R. Gruber // Knowledge Acquisition. – 1993. – V. 5(2). – P. 99-220.
3. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб. : Питер, 2001.
4. Некрашевич С.П. Представление данных в Интернет на основе семантических сетей / С.П. Некрашевич, Д.В. Божко // Искусственный интеллект. – 2006. – № 1. – С. 57-59.
5. Добров Б.В. Вторичное использование лингвистических онтологий: изменение в структуре концептуализации / Б.В. Добров, Н.В. Лукашевич // Восьмая Всероссийская научная конференция «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Владимир-Суздаль, 16-18 октября 2006 г.) – С. 56-64.
6. Буч Г. Язык UML. Руководство пользователя / Буч Г., Рамбо Д., Джекобсон А. – М. : Изд-во ДМК, 2000.

Статья поступила в редакцию 15.04.2009.